

ABSTRACT OF THE DISCLOSURE

The present invention provides a solution to the needs described above through a system and method for efficiently detecting near identities in large DNA databases. The system and method disclosed herein make use of an algorithm used to construct and maintain unique DNA databases wherein the unique database contains no two DNA sequences such that one is nearly identical to a region of the other. The system and method are applicable to problems such as an all against all comparison of all the DNA sequences in a large DNA database, clustering and assembling ESTs into the cDNAs that generated the ESTs, mapping assembled ESTs onto genomic sequence, mapping cDNAs onto genomic sequences and locating alternately spliced cDNAs.